

# Relatório Final do Ensaio de Proficiência de Software - 1ª Rodada Avaliação de Produto de Software



Inmetro  
Instituto Nacional de Metrologia, Qualidade e Tecnologia

**PEP-Inmetro**

Programa de Ensaios de Proficiência do Inmetro

# ENSAIO DE PROFICIÊNCIA DE SOFTWARE

## 1ª RODADA – AVALIAÇÃO DE PRODUTO DE SOFTWARE

Período de Realização: 02/09/19 a 26/11/19

### RELATÓRIO FINAL Nº 008/2019

#### ORGANIZAÇÃO E COORDENAÇÃO PROMOTORA DO ENSAIO DE PROFICIÊNCIA



Instituto Nacional de Metrologia, Qualidade e Tecnologia - Inmetro

Diretoria de Metrologia Científica e Tecnologia - Dimci

Endereço: Av. Nossa Senhora das Graças, 50 – Xerém – Duque de Caxias

RJ – Brasil – CEP: 25250-020

E-mail para contato: [pep-inmetro@inmetro.gov.br](mailto:pep-inmetro@inmetro.gov.br) – Telefone: (21) 2145-3002

#### COMITÊ DE ORGANIZAÇÃO

Adelcio Rena Lemos (Inmetro/Dimci/GT-PEP)

Carla Thereza Coelho (Inmetro/Dimci/GT-PEP)

Jose Ricardo Bardellini da Silva (Inmetro/Dimci/GT-PEP) - Coordenador PEP-Inmetro

Lucas Dias Barros (Inmetro/Dimci/GT-PEP) – Sob supervisão

Paulo Roberto da Fonseca Santos (Inmetro/Dimci/GT-PEP)

Raphael Carlos Santos Machado (Inmetro/Dimci/Dmtic/Lainf)

Rodolfo Saboia Sousa (Inmetro/Dimci/Dmtic)

#### COMITÊ TÉCNICO

Flávia Paiva Agostini (Inmetro/Dimci/Dmtic/Lainf)

Paulo Roberto de Mesquita Nascimento (Inmetro/Dimci/Dmtic/Lainf)

Raphael Carlos Santos Machado (Inmetro/Dimci/Dmtic/Lainf)

Sérgio Câmara (Inmetro/Dimci/Dmtic/Lainf)

Wilson de Souza Melo Junior (Inmetro/Dimci/Dmtic/Lainf)

Wladimir Araújo Chapetta (Inmetro/Dimci/Dmtic/Lainf)

## SUMÁRIO

1. Introdução .....	3
2. Materiais e Métodos .....	4
2.1. Preparação do Item de Ensaio .....	4
2.2. Caracterização, Homogeneidade e Estabilidade do Item de Ensaio .....	4
2.3. Análise Estatística dos Resultados dos Participantes .....	4
2.3.1. Índice de Jaccard .....	4
3. Valores Designados .....	5
4. Resultados e Discussão .....	5
4.1. Caracterização, Homogeneidade e Estabilidade do Item de Ensaio .....	5
4.2. Resultados dos Participantes .....	5
4.2.1. Métodos Utilizados pelos Participantes .....	5
4.2.2. Resultados Reportados pelos Participantes .....	6
4.2.3. Avaliação de Desempenho dos Participantes .....	6
4.2.3.1. Análise objetiva Índice de Jaccard .....	7
4.2.3.2. Discussão dos resultados do Resumo Executivo .....	8
5. Confidencialidade .....	9
6. Conclusões .....	9
7. Participantes .....	10
8. Referências Bibliográficas .....	11

## **1. Introdução**

A avaliação de produtos de software por laboratórios acreditados é uma demanda crescente em todo o mundo. Isso ocorre devido à forte tendência de se utilizar software para automação, gerenciamento, controle e otimização dos mais diversos dispositivos. Deste modo, a avaliação de um produto de software constitui um mecanismo imprescindível para se atestar a conformidade de um produto em relação a um conjunto de requisitos funcionais e não funcionais, garantindo deste modo sua eficiência, confiabilidade e segurança.

Embora Ensaio de Proficiência (EP) sejam realizados em diferentes escopos de avaliação da conformidade, a avaliação de software introduz uma série de novos desafios decorrentes do grau de subjetividade associado à interpretação dos requisitos de um produto de software, bem como sobre os diferentes entendimentos a respeito de como um software pode ser testado [1]. Dessa forma, avaliar a competência de um laboratório na certificação de um produto de software não é uma tarefa trivial. De fato, uma busca nos documentos dos principais organismos acreditadores internacionais mostra que o problema de avaliar proficiência de laboratórios na área de software é ainda um problema em aberto. Ainda assim, dada a importância crescente de tais laboratórios para a realização de avaliações de segurança de software de um número cada vez maior de dispositivos inteligentes e ativos de tecnologia, torna-se premente propor abordagens para iniciar o tratamento de tal problema.

Este relatório descreve os resultados da realização da 1ª rodada de EP na avaliação de laboratórios em escopos relacionados à Avaliação Produto de Software (APS). Este EP se baseou em métricas de cobertura de código como o principal elemento de intercomparação objetiva dos resultados obtidos por cada laboratório. Com a realização do EP coordenado pelo Inmetro, cada laboratório participante teve a oportunidade de avaliar seus métodos e procedimentos de APS e verificar a harmonia com o desempenho dos outros laboratórios.

O presente relatório apresenta os resultados do Ensaio de Proficiência de Software - 1ª Rodada - Avaliação de Produto de Software. Este EP teve como objetivo:

- Determinar o desempenho de laboratórios para o ensaio proposto;
- Contribuir para o aumento da confiança nos resultados de uma APS feita pelos participantes;
- Contribuir para a melhoria contínua das técnicas de APS de cada participante.

## **2. Materiais e Métodos**

### **2.1. Preparação do Item de Ensaio**

O Laboratório de Informática (Lainf) da Divisão de Metrologia em Tecnologias da Informação e Comunicações (Dmtic) do Inmetro foi o responsável pela customização, preparação e disponibilização do item de ensaio. O item de ensaio foi encapsulado dentro do pacote de ensaio, criado por meio do padrão OpenPGP [2], conforme especificados no “Manual de Ensaio de Proficiência de Software”. O próprio Lainf foi o responsável pela disponibilização do pacote de ensaio no ambiente da RNP (<https://filesender.rnp.br>) e informou o link de acesso que habilitou os participantes a acessarem este pacote através do link [www.inmetro.gov.br/metcientifica/ensaio-proficiencia/profiSoftware.asp](http://www.inmetro.gov.br/metcientifica/ensaio-proficiencia/profiSoftware.asp).

O pacote de ensaio foi preparado e entregue pronto para uso, juntamente com instruções operacionais. O pacote teve sua integridade e autenticidade verificadas conforme os procedimentos descritos no “Manual de Ensaio de Proficiência em Software”.

### **2.2 Caracterização, Homogeneidade e Estabilidade do Item de Ensaio**

Para este EP não foram aplicáveis os procedimentos de caracterização, homogeneidade e estabilidade do item de ensaio.

### **2.3 Análise Estatística dos Resultados dos Participantes**

A avaliação de desempenho de cada participante foi baseada na ideia de um objetivo de teste, onde o participante deveria encontrar um conjunto de testes de referência proposto pelo Inmetro, embora esse caso de teste não tenha sido informado aos participantes. Cada participante deveria então proceder com a avaliação do software e tentar modelar um ou mais casos de teste que cobrissem exatamente o mesmo conjunto de linhas cobertas pelos testes de referência. Caso o participante alcançasse esse objetivo, poderíamos concluir que o mesmo teve um bom entendimento da documentação do software, do código relacionado àqueles testes, de como os testes deveriam ser executados – e, por consequência, concluiríamos que o participante seria proficiente na execução de um APS.

#### **2.3.1. Índice de Jaccard**

Uma vez que o objetivo de teste consiste em se encontrar casos de teste que cubram exatamente as mesmas linhas de código cobertas pelos testes de referência, faz-se necessário propor um mecanismo de avaliação que leve em consideração esse aspecto. Em outras palavras, o desempenho do participante será melhor quando seus testes forem mais similares aos testes de referência. Assim, para um dado caso de teste baseado em cobertura de código, denominamos T o conjunto total de linhas do software avaliado, SR o conjunto de linhas cobertas pelos testes de referência e SL o conjunto de linhas cobertas

pelos testes executados pelo participante. Como métrica de similaridade usada para medir o desempenho de cada participante, será adotado o Índice de Jaccard [3], definido a seguir:

$$J(S_R, S_L) = \frac{S_R \cap S_L}{S_R \cup S_L} \quad (1)$$

### 3. Valores Designados

De acordo com os procedimentos disponíveis para o estabelecimento de valores designados pela ABNT NBR ISO/IEC 17043:2011 [4], os valores designados deste EP foram calculados através de métodos estatísticos descritos no item 7.7 da norma ISO 13528:2015 [5], ou seja, valores de consenso de participantes.

A determinação do valor de consenso deste EP foi através da média dos índices  $J(S_R, S_L)$  informados por cada um dos participantes. Assim, o valor  $\bar{J}$  é dado por:

$$\bar{J} = \sum_{i=1}^{labs} \frac{J(S_R, S_L^i)}{labs} \quad (2)$$

Onde *labs* é o número de participantes.

## 4. Resultados e Discussão

### 4.1. Caracterização, Homogeneidade e Estabilidade do Item de Ensaio

Para este EP não foram aplicáveis os procedimentos de caracterização, homogeneidade e estabilidade do item de ensaio.

### 4.2. Resultados dos Participantes

#### 4.2.1. Métodos Utilizados pelos Participantes

Os participantes do EP deveriam utilizar seus próprios métodos de avaliação de software. Basicamente, esperava-se que os métodos fossem compreendidos de pelo menos três etapas características no processo de APS, segundo o escopo de avaliação de software de laboratórios acreditados, a saber:

- 1) Análise da documentação do software: identifica como o software se propõe a satisfazer os requisitos de segurança estabelecidos;
- 2) Inspeção do código fonte do software: verifica se o software foi implementado em conformidade com a documentação analisada;

3) Execução de testes funcionais: confirma se o comportamento dinâmico do software implementado satisfaz a contento os requisitos de segurança estabelecidos.

Adicionalmente, os participantes do EP deveriam utilizar ferramentas providas pelo Lainf para coleta de métricas de teste de software a partir dos testes funcionais realizados. Detalhes sobre tais métricas, bem como sobre as ferramentas providas pelo GT-PEP para esse EP, estavam descritas no “Manual de Ensaio de Proficiência em Software”, disponibilizado site do Inmetro ([www.inmetro.gov.br/metcientifica/ensaio-proficiencia/profiSoftware.asp](http://www.inmetro.gov.br/metcientifica/ensaio-proficiencia/profiSoftware.asp)).

#### 4.2.2. Resultados Reportados pelos Participantes

Todos os participantes reportaram seus resultados conforme os procedimentos estabelecidos no “Manual de Ensaio de Proficiência em Software”. A análise da execução desses procedimentos permite a avaliação objetiva dos participantes, ao mesmo tempo que fornece informações importantes sobre os processos de APS por eles implementados.

Com relação à avaliação objetiva, a mesma seguiu as métricas estabelecidas no Protocolo deste EP, que se baseiam no uso do Índice de Jaccard.

Em complemento à avaliação objetiva, é possível também se discutir dois aspectos relevantes referentes aos processos executados pelos participantes, a saber:

- As informações contidas no Resumo Executivo, que demonstram características da metodologia e ferramentas utilizadas por cada participante;
- A competência demonstrada em atividades envolvendo o uso de ferramentas criptográficas para autenticação e confidencialidade dos resultados gerados por cada participante.

Estes aspectos são discutidos nas subseções a seguir.

#### 4.2.3. Avaliação de Desempenho dos Participantes

A interpretação do desempenho do  $i$ -ésimo laboratório em relação aos demais participantes é associada à comparação de seu Índice de Jaccard  $J_i$ , com o desvio padrão, dado por:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{labs} (\bar{J} - J(S_R, S_L^i))^2}{labs}} \quad (3)$$

O resultado da avaliação do índice  $J_i$  do  $i$ -ésimo participante foi dado pelas seguintes regras:

$J_i \geq \bar{J} - \sigma$	indica desempenho “satisfatório” e não gera sinal;
$\bar{J} - \sigma > J_i \geq \bar{J} - 3\sigma$	indica desempenho “questionável” e gera um sinal de alerta;
$J_i < \bar{J} - 3\sigma$	indica desempenho “insatisfatório” e gera um sinal de ação.

Os resultados dos índices  $J_i$  foram arredondados com duas casas decimais, obedecendo aos critérios de arredondamento.

#### 4.2.3.1. Análise objetiva Índice de Jaccard

A Tabela 1 apresenta os resultados de cada participante, medidos pelo índice de Jaccard. Como é possível observar, todos os participantes obtiveram valor máximo, com índice  $J_i = 1$ . Tal resultado implica que todos os participantes foram capazes de solucionar o desafio de testes referente a esta rodada de EP, e conseqüentemente implementaram casos de teste que cobrem exatamente o conjunto de linhas de código que constituem o desafio proposto.

Tabela 1 - Resultados do índice de Jaccard.

Código do Participante	Índice de Jaccard
01	1,00
05	1,00
07	1,00
12	1,00
14	1,00
18	1,00

Em função dos resultados apresentados na Tabela 1, tem-se que o desvio padrão associado aos valores do índice de Jaccard é zero. Portanto, a aplicação dos critérios estabelecidos indica que todos os participantes apresentaram resultado satisfatório.

O fato de todos os participantes apresentarem resultado idêntico aponta duas questões fundamentais. A primeira é o fato de que a natureza do desafio de teste proposto por esse EP permite que os participantes repitam seus procedimentos de ensaio até obterem um caso de teste que cobre exatamente o conjunto de linhas de código proposto no desafio. A segunda questão é a evidência de que o desafio de testes proposto não apresentou dificuldade suficiente para exercitar a competência técnica dos participantes ao limite.

O aspecto da repetibilidade é natural quando se discute teste de software. Equipes de teste podem repetir exaustivamente determinados procedimentos de verificação enquanto dispõem de tempo para tal. Estabelecendo o cenário descrito como permissa, e traçando um paralelo com o observado nesta rodada de EP, pode-se assumir que o tempo destinado à realização deste EP pelos participantes necessita ser melhor dimensionado nas próximas rodadas. A ideia é que o tempo disponível desde a disponibilização do item de ensaio até o prazo para envio dos resultados pelos participantes seja determinado de forma a tornar o mais restrito desafio da busca pelos casos de teste. Deste modo, será possível avaliar também a capacidade de gestão do participante no aspecto de repetibilidade dos testes por ele propostos.

A questão relacionada à dificuldade dos testes, por sua vez, aponta na direção de que o desafio proposto nessa rodada do EP estava abaixo da competência técnica dos participantes. Tal conclusão é confirmada pelo *feedback* recebido de pelo menos 4 participantes em seus respectivos Resumos Executivos. Eles alegam que os testes realizados foram de dificuldade baixa ou média. Por um lado, esse resultado é positivo por demonstrar que todos os participantes possuem competência minimamente esperada na realização de tarefas imprescindíveis a uma APS. Por outro lado, o mesmo resultado indica que as próximas rodadas do EP devem propor cenários de teste mais desafiadores, de modo a explorar o limite da competência técnica dos participantes.

#### **4.2.3.2. Discussão dos resultados do Resumo Executivo**

Conforme estabelecido no “Manual de Ensaio de Proficiência em Software”, cada participante entregou também, como parte dos resultados do EP, um Resumo Executivo. Tal documento não estava sujeito a qualquer modelo formal predefinido, e nem faz parte da avaliação objetiva dos participantes.

No entanto, seu conteúdo permite avaliar aspectos relevantes da maturidade do processo de APS dos participantes. Os seguintes pontos podem ser destacados:

- Três participantes descrevem as metodologias aplicadas durante o EP, sendo que dois deles apresentam detalhes substanciais da metodologia;
- Dois participantes descrevem o uso de ferramentas de análise complementares àquelas providas no pacote de ensaio, demonstrando como as mesmas foram úteis para a solução do desafio associado ao EP;
- Um dos participantes menciona a análise de modificações realizadas no produto escolhido como item de ensaio, bem como a verificação de CVEs (*Common Vulnerabilities and Exposures* - <https://cve.mitre.org/>), as quais constituem excelentes práticas de avaliação de software;

- Todos os participantes reportam, de forma explícita ou indireta, que os casos de teste propostos no desafio associado a esta rodada de EP eram de complexidade baixa ou média. Um dos participantes inclusive informa um esforço de 6 hh (homem-hora) para solução dos casos de teste, algo bastante abaixo do tempo médio estimado inicialmente pelos proponentes do EP, que seria em torno de 20 hh.

Um último aspecto relevante que deve ser mencionado é a competência demonstrada por todos os participantes em relação ao uso das ferramentas criptográficas necessárias para atendimento às atividades procedimentais deste EP. Tal resultado indireto atesta que os participantes dominam os conceitos associados à troca de informações protegidas por criptografia e ao uso de assinatura de chave pública como mecanismo de verificação de integridade e autenticidade das informações. Tal evidência comprova que os participantes estão aptos a procederem com boas práticas de segurança da informação na proteção de seus registros de ensaio e também dos ativos digitais confiados a eles por seus clientes.

## **5. Confidencialidade**

Cada participante foi identificado por código individual que é conhecido somente pelo próprio participante e pela coordenação do EP. O participante recebeu, via e-mail, o seu código de identificação correspondente à sua participação no EP. Além do código individual, cada participante informou uma chave criptográfica (seguindo um protocolo de criptografia de chave pública) para fins de anonimização e autenticação de seus resultados. Tanto o código individual quanto a assinatura digital correspondente à chave respectiva foram utilizados na identificação do participante e também na geração dos relatórios específicos contendo o registro de resultados. Os resultados poderão ser utilizados em trabalhos e publicações pelo Inmetro respeitando-se a confidencialidade de cada participante.

Conforme estabelecido no item 4.10.4 da ABNT ISO/IEC 17043:2011, em circunstâncias excepcionais, uma autoridade reguladora pode requerer os resultados do EP ao provedor.

## **6. Conclusões**

Todos os participantes apresentaram resultado satisfatório no EP, através da análise de desempenho pelo Índice de Jaccard. O objetivo de determinar o desempenho dos participantes para o ensaio proposto foi atingido. Excepcionalmente, em razão de fatores discutidos na Seção 4.2.3.1, observa-se que todos os participantes obtiveram valor máximo na análise de desempenho, em função de todos demonstrarem capacidade de resolver o desafio de teste de software proposto. Futuras rodadas de EP

levarão em consideração a necessidade de melhor equacionar o tempo disponível para realização das atividades, bem como a proposta de desafios de teste mais complexos, que exercitem ao máximo as competências técnicas dos participantes.

O estabelecimento de ações corretivas e a contínua participação em ensaios de proficiência desta natureza são ferramentas de grande contribuição para o aprimoramento das medições realizadas pelos participantes. Este EP de APS promovido pelo Inmetro, em especial, é uma iniciativa pioneira nesta área de conhecimento, uma vez que, até o presente momento, não há relatos de ensaios de proficiência realizados neste escopo de avaliação.

Finalmente, deve-se ressaltar a importância da participação dos laboratórios em exercícios de EP, por constituir uma ferramenta útil para monitorar os procedimentos de análises usados na rotina e avaliar os resultados de suas medições, tornando-os capazes de desempenhar medições com maior confiabilidade.

## **7. Participantes**

Seis participantes se inscreveram na 1ª Rodada do Proficiência de Software. Todos esses enviaram resultados e tiveram seu desempenho avaliado.

Participaram deste EP laboratórios acreditados no Brasil para o escopo em questão.

A lista dos participantes que enviaram os resultados à coordenação desse EP é apresentada na Tabela 2. É importante ressaltar que a numeração da tabela é apenas indicativa do número de participantes no EP, não estando, em hipótese alguma, associada à sua identificação na apresentação dos resultados.

**Tabela 2 – Participantes**

<b>Organização</b>	
1.	Fundação CPQD
2.	Fundação Instituto Nacional de Telecomunicações
3.	Green Hat Segurança da Informação Ltda.
4.	Instituto de Ciências Matemáticas e de Computação
5.	UBEA - União Brasileira de Assistência e Educação LABELO-PUCRS – Laboratórios Especializados em Eletroeletrônica

Organização	
6.	Universidade Federal do Rio de Janeiro - LASPI

Total de participantes: 6.

## 8. Referências Bibliográficas

- [1] MYERS, Glenford J.; SANDLER, Corey; BADGETT, Tom. “The art of software testing”. John Wiley & Sons, 2011.
- [2] STALLINGS, W. e BROWN, L. Computer Security: Principles and Practice. 4ª Edição, Pearson, 2018.
- [3] JACCARD, Paul. The distribution of the flora in the alpine zone. 1. New Phytologist, v. 11, n. 2, p. 37-50, 1912.
- [4] ABNT NBR ISO/IEC 17043, Avaliação de conformidade - Requisitos gerais para ensaios de proficiência, ABNT, Rio de Janeiro, 2011.
- [5] ISO 13528, Statistical methods for use in proficiency testing by interlaboratory comparison, ISO, Geneva, 2015.
-



Programa de Ensaio da Proficiência do Instituto Nacional de Metrologia, Qualidade e Tecnologia - PEP-Inmetro  
Av. Nossa Senhora das Graças, 50 - Xerém - Duque de Caxias - RJ - Brasil CEP: 25250-020  
Tel/Fax: +55 21 2679-9745 - [www.inmetro.gov.br](http://www.inmetro.gov.br) - E-mail: [pep-inmetro@inmetro.gov.br](mailto:pep-inmetro@inmetro.gov.br)